

O presente trabalho se insere no projeto de pesquisa bibliométrica realizado no Departamento de Ciência da Informação da FABICO/UFRGS, intitulado a Ciência no Rio Grande do Sul: mapeamento da produção e colaboração nos anos de 2000 a 2009. Trata-se de um estudo avaliativo dos procedimentos técnicos efetuados para a limpeza dos dados extraídos da base de dados Web of Science (WoS) para o projeto de pesquisa em andamento, que contribui para evidenciar a necessidade da padronização nos dados disponibilizados nas bases de dados. Consiste nos processos de identificação de autoria, co-autoria e filiação dos trabalhos publicados por autores vinculados a pelo menos uma instituição sediada no Rio Grande do Sul e que estão indexados na WoS. A partir da recuperação e da organização dos dados bibliográficos, foram contabilizados 49.019 nomes de autores e co-autores. Partindo da classificação dos nomes em ordem alfabética, observa-se que muitas vezes um mesmo autor é identificado de diferentes formas. Como não há padronização desses dados na WoS, os registros recuperados contém diversas imperfeições, dentre estas, erros de grafia e ausência de componentes essenciais para a completa identificação dos autores, como por exemplo, o registro completo de sobrenomes. Estes erros necessitam ser corrigidos manualmente, consumindo tempo na função de padronização dos dados da pesquisa. A fim de garantir a consistência dos dados, os Currículos Lattes dos autores foram consultados para a identificação de autoria. Quando mesmo essas informações são insuficientes, os dados contidos no Currículo Lattes de um dos co-autores do trabalho em questão fornecem as informações necessárias, ou seja, sua rede de colaboração é a ferramenta utilizada para identificá-lo. Após efetuar a análise descritiva dos autores da pesquisa, observou-se que aproximadamente 10 % do número total de nomes de autores indexados pela WoS apresentavam variações em sua forma de apresentação. Este fato, além de afetar a qualidade da recuperação da informação, também influencia na composição dos índices bibliométricos e, desta forma, pode levar o pesquisador a conclusões equivocadas. Autores que utilizam diferentes formas para a apresentação de seus nomes provocam a dispersão do número total de registros de autoria, acarretando na dificuldade da avaliação da produção científica individual e, conseqüentemente, institucional, regional e nacional. Sugere-se a adoção de formas padronizadas de nome de autor na produção bibliográfica.